An EM Framework for Segmentation of Tissue Mixtures from Medical Images

Zhengrong Liang[1,2], Xiang Li[1], Daria Eremina[3], and Lihong Li[1]
Departments of Radiology[1], Computer Science[2] and Applied mathematics[3], State University of New York
Stony Brook, NY 11794, USA

Abstract—**Image segmentation plays a major role in quantitative image analysis and computer aided detection (CAD) and diagnosis (CADx) for clinical applications. Conventional segmentation assigns a single label to each voxel, neglecting the partial volume (PV) effect. This work presents an EM (Expectation Maximization) framework for segmentation of tissue mixture in each voxel. Image data and tissue mixture models, EM algorithm for mixture quantification, prior model for regularization on the mixtures, and multi-spectral MR (magnetic resonance) data characterization are described in details. Preliminary results from CT (computed tomography) and MR images are reported to demonstrate its potential for clinical use.**

Keywords—**Image segmentation, tissue mixture, maximum *a posteriori* (MAP) probability, EM algorithm**

I. Introduction

Image segmentation is a major component of image processing methodology, facilitating quantitative analysis and visualization of clinical features in medical images toward diagnosis, treatment/surgical planning and follow-up evaluation. Conventional image segmentation assigns a single label to reflect a specific tissue type in each voxel and does not provide the percentage of each tissue type in that voxel. This can result in a significant error in quantitative image analysis [1]. A logical solution is to determine each tissue percentage in each voxel. Quantifying the tissue mixture in each voxel has been attempted in the recent years [2] with a noticeable success. This work presents a framework for mixture segmentation based on the well-established EM (Expectation Maximization) algorithm [3].

II. Methodology

A. Image Data Model

Given an acquired image $\mathbf{Y} = \{Y_i\}$, i = 1,2,…,I, over I voxels. Each voxel value Y_i is an observation of a random process around a mean $\overline{Y}_i$ and variance s_i^2, i.e.,

$$Y_i = \overline{Y}_i + n_i \tag{1}$$

where n_i reflects the associated noise with zero mean and variance of s_i^2. Assume that the noise follows a Gaussian distribution, then

$$P(Y_i \mid q_i) = \frac{1}{\sqrt{2p\, s_i^2}} e^{-(Y_i - \overline{Y}_i)^2 / 2 s_i^2} \tag{2}$$

where q_i reflects the model parameters of mean and variance $\{\overline{Y}_i, s_i^2\}$. Assume that all voxel values $\{Y_i\}$ are statistically independent from each other, given the mean distribution $\{\overline{Y}_i\}$, then we have

$$P(\mathbf{Y} \mid \Theta) = \prod_{i=1}^{I} P(Y_i \mid q_i) \tag{3}$$

where $\Theta = \{q_i\}$, i = 1,2,…, I.

It is noted that given the mean $\overline{Y}_i$ at voxel i, repeated observations at that voxel will render the variance s_i^2. Therefore, the probability density function in equations (2) and (3) is defined for the given model parameters, which can be utilized to assign an appropriate label for each voxel, given the observed image $\{Y_i\}$ [4]. However, the model parameters describe only the global properties of each voxel as a whole volume and do not consider its substructures or mixture. In many medical imaging applications, each voxel can have more than one tissue types due to the limited spatial resolution. Ignoring the substructures will result in the well-known partial volume (PV) effect. Consideration of the substructures is given below.

B. Tissue Mixture Model

The acquired image $\{Y_i\}$ reflects K tissue types distributing inside the body. Within each voxel volume, there possibly are K tissue types, where each tissue type has a contribution to the observed voxel value Y_i at voxel i. Let X_{ik} be the contribution of tissue type k to the observation Y_i. It is clear that X_{ik} is also a random process around a mean $\overline{X}_{ik}$ and variance s_{ik}^2, i.e.,

$$X_{ik} = \overline{X}_{ik} + n_{ik} \tag{4}$$

where n_{ik} reflects the associated noise with zero mean and variance of s_{ik}^2. Assume that the noise n_{ik} also follows a Gaussian distribution, then

$$P(X_{ik} \mid q_{ik}) = \frac{1}{\sqrt{2p\, s_{ik}^2}} e^{-(X_{ik} - \overline{X}_{ik})^2 / 2 s_{ik}^2} \tag{5}$$

where q_{ik} reflects the model parameters of mean and variance $\{\overline{X}_{ik}, s_{ik}^2\}$. Assume that all K tissue types have no correlation, then

$$P(\hat{X}_i \mid \hat{q}_i) = \prod_{k=1}^{K} P(X_{ik} \mid q_{ik}) \tag{6}$$

where $\hat{X}_i = \{X_{ik}\}$ and $\hat{q}_i = \{q_{ik}\}$, k = 1,2,…K, are vectors of length K for voxel i. By the assumption that all voxel variables $\{\hat{X}_i\}$, i = 1,2,…,I, are statistically independent, given the model parameters, then we have

$$P(\mathbf{X} \mid \Theta) = \prod_{i,k=1}^{I,K} P(X_{ik} \mid q_{ik}) \tag{7}$$

where $\mathbf{X} = \{\hat{X}_i\}$, and parameter $\Theta = \{\hat{q}_i\}$ has the same meanings as that as defined in equation (3). The relation between $\{q_i\}$ in equation (3) and $\{\hat{q}_i\}$ in equation (6) can be derived from the relation of the observed datum Y_i and its components from different tissue types in voxel *i*, i.e.,

$$Y_i = \sum_{k=1}^{K} X_{ik} \tag{8}$$

which results in

$$\overline{Y}_i = \sum_{k=1}^{K} \overline{X}_{ik}\ , \text{and}\ n_i = \sum_{k=1}^{K} n_{ik}\ \text{ or }\ s_i^2 = \sum_{k=1}^{K} s_{ik}^2\ . \tag{9}$$

Let voxel *i* be partitioned into *K* compartments according to the tissue means of $\{\overline{X}_{ik}\}$, *k* = 1,2,...*K*, and Z_{ik} be the fraction of tissue type *k* in voxel *i*, which satisfies

$$\sum_{k=1}^{K} Z_{ik} = 1. \tag{10}$$

Furthermore, let m_k and n_k be the mean and variance of tissue type *k* fully filling a voxel. Then we have

$$\overline{X}_{ik} = Z_{ik} m_k\ ,\quad \overline{Y}_i = \sum_{k=1}^{K} Z_{ik} m_k$$
$$s_{ik}^2 = Z_{ik} n_k\ ,\ \text{ and }\ \ s_i^2 = \sum_{k=1}^{K} Z_{ik} n_k\ . \tag{11}$$

The tissue model parameters $\{m_k, n_k\}$ reflect the inherent properties of tissue type *k* in the observed image $\{Y_i\}$. When a voxel is fully filled by tissue type *k*, where the mixture parameter Z_{ik} equals to 1, then the observed datum Y_i equals to the tissue contribution X_{ik} at that voxel. This is expected because both the image data and tissue mixture models must be consistent. The mixture parameters $\{Z_{ik}\}$ completely specify the tissue distribution in the body and, therefore, are the target for quantitative image analysis.

Estimation of the mixture parameters $\{Z_{ik}\}$ for more than one tissue types from an acquired image $\{Y_i\}$ is a problem of missing data, as reflected by equation (8) as a mapping from more than one (right) to one (left) variables. For *K* = 2 (i.e., two tissue types), we have *I* measurements $\{Y_i\}$ to estimate *I* mixture parameters $\{Z_{ik}\}$ with the constraint of equation (10), in addition to the four tissue model parameters $\{m_k, n_k\}$. A well-established strategy for solving this missing data problem is the EM algorithm.

C. EM Algorithm for Mixture Quantification

By the EM terminology, Y_i is an observable random variable and incomplete (in reflecting the underline true tissue information), while X_{ik} is an unobservable random variable and reflects the complete information for each underline tissue type. The sampling densities for these two random variables are related by

$$P(\mathbf{Y} \mid \Theta) = \int_{\{Y_i = \sum_{k=1}^{K} X_{ik}\}} P(\mathbf{X} \mid \Theta)\, dX \tag{12}$$

where the integral is over all possible configurations of $\{X_{ik}\}$ under the constraint of equation (8).

The EM algorithm seeks the maximum likelihood (ML) solution for the model parameters Θ (including the mixture and tissue model parameters) via the complete sampling density by interleaved Expectation and Maximization steps in an iterative manner.

The E-step computes the conditional complete-sampling density, given the observed data $\mathbf{Y}$ and the *n*-th iterated estimate of the mixture and tissue model parameters $\Theta^{(n)}$,

$$Q(\Theta \mid \Theta^{(n)}) = \mathrm{E}[\ln P(\mathbf{X} \mid \Theta) \mid \mathbf{Y}, \Theta^{(n)}] = -\tfrac{1}{2} IK \ln(2p) - (\tfrac{1}{2}) \times \sum_{ik} \left[\ln(Z_{ik} n_k) + \tfrac{1}{Z_{ik} n_k}[(X_{ik}^2)^{(n)} - 2X_{ik}^{(n)} Z_{ik} m_k + Z_{ik}^2 m_k^2]\right] \tag{13}$$

where the conditional means $X_{ik}^{(n)}$ and $(X_{ik}^2)^{(n)}$ for X_{ik} and X_{ik}^2 respectively are given by

$$X_{ik}^{(n)} = \mathrm{E}[X_{ik} \mid Y_i, \Theta^{(n)}] = Z_{ik}^{(n)} m_k^{(n)} + \frac{Z_{ik}^{(n)} n_k^{(n)}}{\sum_{j=1}^{K} Z_{ij}^{(n)} n_j^{(n)}} \cdot (Y_i - \sum_{j=1}^{K} Z_{ij}^{(n)} m_j^{(n)}) \tag{14}$$

$$(X_{ik}^2)^{(n)} = \mathrm{E}[X_{ik}^2 \mid Y_i, \Theta^{(n)}] = (X_{ik}^{(n)})^2 + Z_{ik}^{(n)} n_k^{(n)} \cdot \frac{\sum_{j \neq k}^{K} Z_{ij}^{(n)} n_j^{(n)}}{\sum_{j=1}^{K} Z_{ij}^{(n)} n_j^{(n)}} \tag{15}$$

where $(X_{ik}^{(n)})^2$ is the square of the *n*-th iteration of $X_{ik}^{(n)}$.

The M-step determines the (*n*+1)-th iterated estimate, which maximizes the conditional complete-sampling density of equation (13). For the mean parameter m_k, we have

$$\frac{\partial Q(\Theta \mid \Theta^{(n)})}{\partial m_k}\Big|_{(n+1)} = \sum_i \frac{1}{n_k} (X_{ik}^{(n)} - Z_{ik} m_k)\Big|_{(n+1)} = 0\ , \tag{16}$$

which results in

$$m_k^{(n+1)} = \frac{\sum_i X_{ik}^{(n)}}{\sum_i Z_{ik}^{(n)}}\ , \tag{17}$$

i.e., the mean for tissue type *k* is the summation of its contributions divided by the summation of its fractions over all voxels. It concurs with our expectation. For the variance parameter n_k, we have

$$\frac{\partial Q(\Theta \mid \Theta^{(n)})}{\partial n_k}\Big|_{(n+1)} = \sum_i \{\frac{1}{n_k} - \frac{1}{Z_{ik} n_k^2}[(X_{ik}^2)^{(n)} - 2X_{ik}^{(n)} Z_{ik} m_k + Z_{ik}^2 m_k^2]\}\Big|_{(n+1)} = 0 \tag{18}$$

which results in

$$n_k^{(n+1)} = \frac{1}{I} \sum_i \frac{(X_{ik}^2)^{(n)} - 2X_{ik}^{(n)} Z_{ik}^{(n)} m_k^{(n)} + (Z_{ik}^{(n)} m_k^{(n)})^2}{Z_{ik}^{(n)}}\ . \tag{19}$$

Its meanings can be seen for a special case of single tissue type, where $Z_{ik} = 1$ and $(X_{ik}^2)^{(n)} = (X_{ik}^{(n)})^2$, the variance for that tissue type equals the summation of all voxel's

variances divided by the number of all the voxels. It concurs again with our expectation.

For the mixture parameter Z_{ik}, the constraint of equation (10) must be satisfied. Therefore, the $(n+1)$-th iterated estimate $Z_{ik}^{(n+1)}$ is given by minimizing $Q(.)$ subject to the constraint of equation (10).

For cases of two tissue types, i.e., k = 1 and 2, the E-step of equation (13) becomes determination of

$$Q(.) = -(\frac{1}{2})\sum_i \{\ln(Z_{i1}n_1) + \frac{(X_{i1}^2)^{(n)}}{Z_{i1}n_1} - \frac{2X_{i1}^{(n)}m_1}{n_1} + \frac{Z_{i1}m_1^2}{n_1} + \ln[(1-Z_{i1})n_1] + \frac{(X_{i2}^2)^{(n)}}{(1-Z_{i1})n_2} - \frac{2X_{i2}^{(n)}m_2}{n_2} + \frac{(1-Z_{i1})m_2^2}{n_2}\} - \frac{I}{2}\ln(2p) \quad (20)$$

where $Z_{i2} = 1 - Z_{i1}$ was used. Maximization of $Q(.)$ with respect to $Z_{i1} = x$ is given by setting the partial differentiation to be zero, i.e.,

$$a \cdot x^4 - 2 \cdot a \cdot x^3 + b \cdot x^2 + c \cdot x - d = 0 \quad (21)$$

where

$$a = n_2^{(n)}(m_1^{(n)})^2 - n_1^{(n)}(m_2^{(n)})^2$$
$$b = a + n_1^{(n)}(X_{i2}^2)^{(n)} - n_2^{(n)}(X_{i1}^2)^{(n)} - n_1^{(n)}n_2^{(n)}$$
$$c = n_1^{(n)}n_2^{(n)} + 2n_2^{(n)}(X_{i1}^2)^{(n)}, \text{ and } d = n_2^{(n)}(X_{i1}^2)^{(n)} \quad (22)$$

Solving equation (21) for $x = Z_{i1}^{(n+1)}$, then we have $Z_{i2}^{(n+1)} = 1 - Z_{i1}^{(n+1)}$.

The ML-EM estimation of I+4 parameters from I measurements can be improved with consideration of valid *a priori* information for a maximum *a posteriori* (MAP-EM) estimation.

D. Prior Model for Mixture Regularization

The EM model parameter estimation above can be regularized by imposing penalty on each tissue type (under the above assumption that all K tissue types are not correlated). The penalty on each tissue type is imposed upon the mixture parameter $\{Z_{ik}\}$ in our case.

By a Gibbs model on a Markov random field (MRF) framework, the penalty on the mixture parameter $\{Z_{ik}\}$ has the form of

$$P(Z_{ik} \mid \{Z_{ike_i}\}) = \frac{1}{a}e^{-b \cdot U(Z_{ik} - Z_{ike_i})} \quad (23)$$

where $\{Z_{ike_i}\}$ are the neighbors of Z_{ik} surrounding voxel i, a is a normalization constant and b is an adjustable parameter controlling the degree of the penalty. The potential function $U(.)$ for a quadratic form can be written as

$$U(.) = \sum_{r \in e_i} w_{ir} \cdot (Z_{ik} - Z_{ike_i})^2 \quad (24)$$

where w_{ir} is a weighting factor for different orders of neighbors. By including this penalty term into equation (13) for a penalized ML-EM or MAP-EM parameter estimation, we have

$$Q(\Theta \mid \Theta^{(n)}) = \mathrm{E}[\ln\{P(\mathbf{X} \mid \Theta)P(\Theta \mid \Theta_e)\} \mid \mathbf{Y}, \Theta^{(n)}] = C(.) - (\tfrac{1}{2}) \times \sum_k \left[\ln(Z_{ik}n_k) + \tfrac{1}{Z_{ik}n_k}[(X_{ik}^2)^{(n)} - 2X_{ik}^{(n)}Z_{ik}m_k + Z_{ik}^2m_k^2] + 2bU(.)\right] \quad (25)$$

where Θ_e represents a neighborhood system and $C(.)$ is a constant for maximizing function $Q(.)$. Maximizing $Q(.)$ for m_k, n_k, and Z_{ik} were described above.

As discussed above, from an acquired single-channel image $\{Y_i\}$, accurate estimation of the mixture parameters for more than two tissue types is very challenging due to the severe missing of measurements. Although drawing more samples from a region of similar densities Y_i from the image can improve the estimation [5], multi-channel or -spectral measurements provide sufficient information for unbiased statistically reliable estimate of mixtures in each voxel.

E. Multi-spectral Data Characterization

In MR (magnetic resonance) imaging, a multi-spectral image series can be acquired in real time as T_1, T_2, and proton density P_D weighted. For each image, the observed datum at voxel i is specified by equations (1) and (2), and for all the voxels, the observation is specified by equation (3). The K tissue types in a single image are characterized by equations (4)-(11). However, the fraction of tissue type k at voxel i for the three images is defined by the same mixture parameter Z_{ik}, although the contributions of a tissue fraction to the three observed images have different means and variances. Let t indicate a single image in the image series consisting of T_1, T_2, and P_D weighted scans, the relations in equation (11) become

$$\overline{X}_{ikt} = Z_{ik}m_{kt}, \quad \overline{Y}_{it} = \sum_{k=1}^{K} Z_{ik}m_{kt}$$
$$s_{ikt}^2 = Z_{ik}n_{kt}, \text{ and } s_{it}^2 = \sum_{k=1}^{K} Z_{ik}n_{kt}. \quad (26)$$

Since they are acquired at different times, the three images in the multi-spectral T_1, T_2, and P_D series are statistically independent. Translating this independence into the unobservable random variable $\mathbf{X}$, equation (7) becomes

$$P(\mathbf{X} \mid \Theta) = \prod_{i,k,t=1}^{I,K,3} P(X_{ikt} \mid q_{ikt}) \quad (27)$$

where t = 1,2,3 and indicates the T_1, T_2 and P_D weighted images.

The E-step of the EM algorithm for the multi-spectral images computes, from equation (13), the conditional density of

$$Q(\Theta \mid \Theta^{(n)}) = \mathrm{E}[\ln P(\mathbf{X} \mid \Theta) \mid \mathbf{Y}, \Theta^{(n)}] = -(\tfrac{3}{2})IK\ln(2p) - (\tfrac{1}{2}) \times \sum_{ikt}\left[\ln(Z_{ik}n_{kt}) + \tfrac{1}{Z_{ik}n_{kt}}[(X_{ikt}^2)^{(n)} - 2X_{ikt}^{(n)}Z_{ik}m_{kt} + Z_{ik}^2m_{kt}^2)\right] \quad (28)$$

where the conditional means for $X_{ikt}^{(n)}$ and $(X_{ikt}^2)^{(n)}$ are given by

$$X_{ikt}^{(n)} = \mathrm{E}[X_{ikt} \mid \{Y_{it}\}, \Theta^{(n)}]$$
$$= Z_{ik}^{(n)} m_{kt}^{(n)} + \frac{Z_{ik}^{(n)} n_{kt}^{(n)}}{\sum_{j=1}^{K} Z_{ij}^{(n)} n_{jt}^{(n)}} \cdot (Y_{it} - \sum_{j=1}^{K} Z_{ij}^{(n)} m_{jt}^{(n)}) \quad (29)$$

$$(X_{ikt}^2)^{(n)} = \mathrm{E}[X_{ikt}^2 \mid \{Y_{it}\}, \Theta^{(n)}]$$
$$= (X_{ikt}^{(n)})^2 + Z_{ik}^{(n)} n_{kt}^{(n)} \cdot \frac{\sum_{j \neq k}^{K} Z_{ij}^{(n)} n_{jt}^{(n)}}{\sum_{j=1}^{K} Z_{ij}^{(n)} n_{jt}^{(n)}} . \quad (30)$$

The M-step for estimate of the mean parameter is given by maximizing $Q(.)$ with respect to m_{kt}. The result is

$$m_{kt}^{(n+1)} = \frac{\sum_i X_{ikt}^{(n)}}{\sum_i Z_{ik}^{(n)}} . \quad (31)$$

The estimate of the variance parameter s_{ikt}^2 is given by

$$n_{kt}^{(n+1)} = \frac{1}{I} \sum_i \frac{(X_{ikt}^2)^{(n)} - 2X_{ikt}^{(n)} Z_{ik}^{(n)} m_{kt}^{(n)} + (Z_{ik}^{(n)} m_{kt}^{(n)})^2}{Z_{ik}^{(n)}} . \quad (32)$$

The M-step for estimate of the mixture parameter is given by maximizing $Q(.)$ with respect to Z_{ik} with the constraint of equation (10). For more than two tissue types, there may not have a simple formula for the $(n+1)$-th iterated solution. This complexity can be avoid by maximizing the probability of the observed data $\mathbf{Y} = \{Y_{it}\}$, instead of working on the probability of the unobserved data $\mathbf{X} = \{X_{ikt}\}$, i.e., by minimizing the following linear functional with the constraint of equation (10)

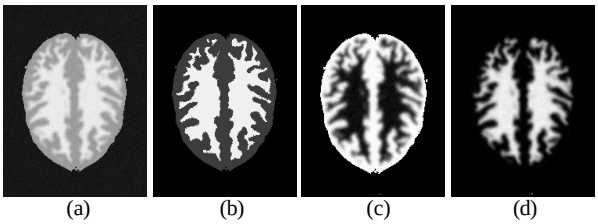
$$F = \frac{1}{2} [Y_i - m^T Z_i]^T \Sigma_i^{-1} [Y_i - m^T Z_i] \quad (33)$$

where equations (2) and (9) have been used for the multi-spectral data, and Σ_i is the covariance matrix with s_{ikt}^2 being its diagonal elements.

By the multi-spectral series of three images of T_1, T_2, and P_D weighted, we can estimate reliably the mixture parameters up to 4 tissue types, which may be the white matter, gray matter, cerebral spinal fluid (CSF) and an abnormality class. With more than 4 tissue types, we may need to acquire more images, such as diffusion image, in the series or include more restricted constraints.

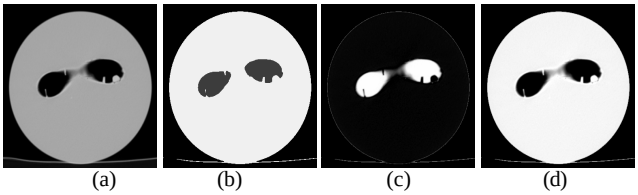
III. Results

A digitized MR image was simulated, in which each voxel was classified as a unique tissue type. The total number of classed was set up to 2 for simplicity. The means of these two classes are 50 and 80, respectively. A low-pass filter was applied to each class image. The filtered class images were added together to mimic the PV effect. White Gaussian noise was added to the smoothed image with PV effect, as shown in Figure 1(a). We adopted a conventional MAP-MRF label segmentation method [4] and compared its performance with this presented method on the simulated image. It is clearly seen that the label segmentation of Figure 1(b) misses a noticeable details on the boundaries. Compared with the label segmentation, the mixture result represents a more accurate anatomic structure, as shown in Figures 1(c) and 1(e). We also applied our method to experimental colon phantom CT (computed tomography) image. Figure 2 compares the results of label and mixture segmentations.



(a) (b) (c) (d)

Figure 1. Simulation results: (a) simulated image, (b) label segmentation, (c) mixture segmentation for class 1, (d) mixture segmentation for class 2.



(a) (b) (c) (d)

Figure 2. Experimental study: (a) colon phantom CT image, (b) label segmentation, (c) mixture segmentation for air, (d) mixture segmentation for soft tissue.

IV. Conclusion

The derivations are mathematically exact for the EM framework, based on the corresponding assumptions. The preliminary results by both simulated and experimental image datasets are very encouraging.

Acknowledgment

This work was supported in part by a NIH grant #CA82402 and #HL51466.

References

[1] H. Choi, D. Haynor, and Y. Kim, "Partial volume tissue classification of Multichannel magnev. 10, pp. 395-407, Sept, 1994.

[2] K. Leemput, F. Maes, D. Vandermeulen, and P. Suetens, "A unifying framework for partial volume segmentation of brain MR images," *IEEE Trans. Med.* Imaging, vol. 22, pp. 105-119, Jan, 2003.

[3] A. Dempster, N. Laird, and D. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *J R Stat. Soc*., vol. 39B, pp. 1-38, 1977.

[4] Z. Liang, J. MacFall, and D. Harrington, "Parameter Estimation and Tissue Segmentation from Multi-spectral MR Images," *IEEE Trans. Med. Imaging*, vol. 13, pp. 441-449, 1994.

[5] Z. Liang, R. Jaszczak and E. Coleman, "Parameter Estimation of Finite Mixtures Using the EM Algorithm and Information Criteria with Application to Medical Image Processing," *IEEE Trans. Nucl. Science*, vol. 39, pp. 1126-1133, 1992.